TECHNOLOGY ADVICE YOU CAN TRUST

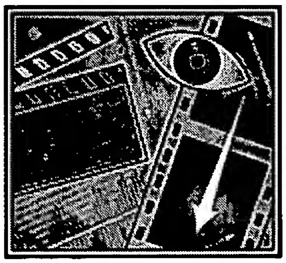
PC WORLD

Topics > Components > Boards > Graphics Boards >

How It Works: Graphics Boards

How today's graphics boards deliver that 3D image from your PC to your screen.

Joel Strauch
Monday, October 09, 2000



Graphics Board: A system component, usually an expansion card, that produces 2D or 3D images on a CRT monitor.

As one of the most important parts of your PC (other than the CPU itself), the graphics board translates the binary ones and zeros of computing into the images that we can interact with on our monitors. Simply put, we couldn't use computers in many of the ways we now do without advanced graphics technology.

Here are the vital stats:

- Graphics boards handle all 2D and 3D calculations and rendering, offloading these intensive tasks from the CPU.
- State-of-the-art graphics boards cost around $300, though you can find reasonably speedy 3D performers for about half that price.
- New graphics chip sets increase graphics board performance, and vendors release new chip sets about every 6 to 12 months.
- Nearly all modern graphics boards connect to the Accelerated Graphics Port slot in a PC and sport at least 16MB of onboard memory.
- Only the most advanced 3D games and high-end computer-aided design come close to taxing modern graphics boards' abilities.

Advertisement

Images you see on your monitor take a complex route from inside the PC. When the application you're running wants to create an image, it sends a request for help to the part of the operating system that connects to the graphics board (called the *graphics driver interface).* In response, the graphics driver--software that acts as a go-between for the OS and the graphics board--listens to the instructions from either the OS or the application, then takes the digital data and converts it into a format that the graphics board can understand.

Next, the driver pipes the newly formatted digital data to the graphics board for rendering. If

you have a PC made after 1998, the data travels to the board via a slot on the motherboard called the Accelerated Graphics Port. (Older PCs don't have this AGP slot, so their graphics boards plug into a standard PCI slot.)

The data's first stop, once pumped into the board, is the temporary storage space of memory, either on the board itself or inside system memory. Then the board's processor, called the *graphics processing unit* (or GPU), turns the digital data into *pixels,* the sets of colored dots that make up any image you see on the monitor. The volume of pixels produced by the board is enormous: When your screen resolution is set to 1024 by 768, the graphics board calculates the precise color for, and produces, data for 786,432 pixels to draw the screen---and it repeats this process 30 to 90 times each second.

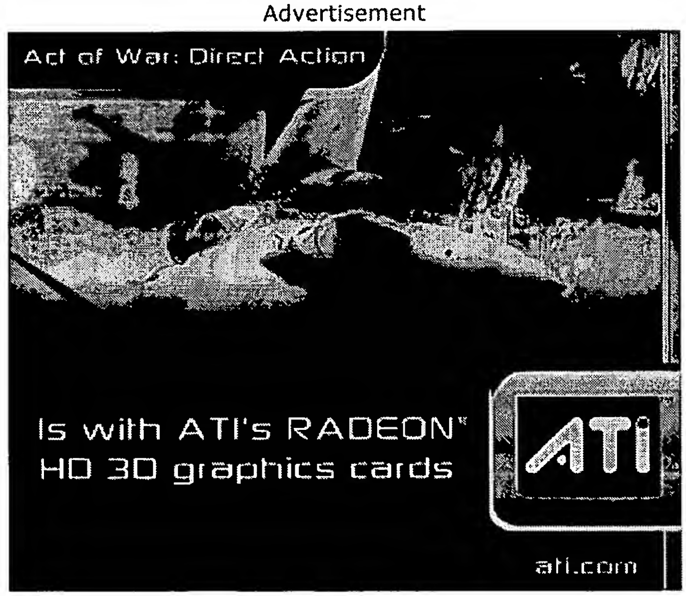
Related Topics: CRT, Graphics & Multimedia, Software, Technology

Rocketing Data to the Monitor

At this point, the pixels aren't ready to paint the screen. After each chunk of image data is converted to pixel data, it gets sent back to memory for storage. Most CRT monitors comprehend only analog signals. That's where the graphics board's random access memory digital-to-analog converter, or RAMDAC, comes into play. As its name implies, the RAMDAC changes the digital data for each pixel into the red, green, and blue analog signals that your monitor can use to display the image. The faster the RAMDAC is able to convert data (measured in megahertz), the higher the resolution the graphics board can produce, and the faster it can refresh. (A few CRT--and many LCD displays can accept the digital signal; those displays require a special card with no RAMDAC.)

The monitor displays analog pixels one line at a time, and to reproduce the illusion of movement, the graphics board must produce many screenfuls of pixels every second. Graphics board vendors call the speed at which the board paints pixels its *fill rate;* current boards provide fill rates in the range of a gigapixel (roughly 1 billion pixels) per second. Faster fill rates lead directly to faster *frame* (or *screen refresh*) *rates*--how many times the entire screen can be redrawn each second. Rates of at least 30 frames per second (the frame rate for television and video) trick the eye into seeing fluid motion.

Advertisement



When you raise the resolution (a lot of cards support resolutions of up to 2048 by 1536 pixels), you add to the graphics board's workload by forcing it to produce a larger number of pixels. Likewise, when you increase the color depth from 16-bit "high color" to 32-bit "true color," it adds to the complexity of the work the card must perform. Because a larger frame requires more pixels to fill it, the frame rate can slow down if the workload reaches the maximum capacity of the GPU and RAMDAC to produce pixel data.

Meeting the 3D Challenge

Like Superman bending steel, newer graphics boards easily manage the processing of two-dimensional images (such as those created when you browse the Web or work in office applications). These so-called 2D images get rendered only in the vertical and horizontal, like drawing on a piece of paper.

But to create realistic 3D images on a two-dimensional surface like your monitor requires more out of the card: the perception of depth. If colored pixels or dots are the building blocks of 2D graphics, polygons--shaded in with textures on the sides to add the illusion of depth--make up 3D images. Most people encounter 3D images when they play games on their PC, but a handful of applications, from CAD to 3D modeling programs, also use the 3D processing power in the card. Handling polygons takes far more processing power than assigning colors to pixels, but today's GPUs--some of which have as much processing power as the CPUs in budget PCs--are up to the task.

Tricks of the 3D Trade

Graphics boards use the large bag of tricks built into their firmware and drivers to render 3D perspective in two dimensions, tricking the eye into believing that polygons displayed on a flat monitor tube have depth. Depending on your board, some of these features are handled by the GPU, while others are handled by your computer's CPU through the graphics driver. The tricks include the following:

- **Hardware transform and lighting** handles the placement of polygons and lighting effects.
- **Bump mapping** applies the look of smoothness or roughness to polygon textures, which helps add to the illusion of depth.
- **Antialiasing** removes the jagged edges of diagonal lines in drawn images.
- **MIP mapping** prevents the card from trying to draw detailed polygon textures on "distant" objects (like the ground in flight simulator games, when you're flying at cruising altitude) until you get closer, which saves processing power.

In addition to the graphics driver, software called the Application Programming Interface controls all the other aspects of this 3D rendering process in a common language that the graphics board (or its driver) can understand. There are several competing APIs, including Microsoft's Direct3D (part of DirectX), OpenGL, and the proprietary GLide API made by 3dfx for its own boards. Most graphics boards are designed to work with more than one API for their 3D rendering.

Related Topics: CRT, Graphics & Multimedia, Software, Technology

Which Graphics Board Is Right for You?

Since the earliest days of PC-compatible computers, PC makers have used graphics boards to speed up the display process. The earliest PCs used graphics boards to take the load off the CPU when it had to render each character in the text-based DOS environment. With the advent of modern GUI operating systems like Windows--and 3D games--rendering got more complex, and

graphics boards became much more critical.

As the graphics board market matured, 3dfx, with its Voodoo line of graphics accelerators, quickly became king of the 3D graphics board market. But over time, other companies began to challenge 3dfx's market dominance: NVidia, with its GeForce line (currently the only major line of graphics chips licensed to multiple manufacturers), has wrested the throne away from the erstwhile monarch.

Advertisement

While the two giants still battle it out for supremacy in the graphics board arena, other companies have managed to chip out a handhold, as well. ATI, always a stalwart competitor in the budget graphics board market, recently stepped into the high-end ring with a new line of boards powered by a GPU it calls Radeon. Led by NVidia, the graphics chip set development cycle for the entire industry has sped up to six months, with each vendor releasing a major new GPU chip set at least once a year. Other companies such as Matrox,SiS, and S3 compete in the medium- to low-end graphics board market.

Most of today's graphics boards use at least 32MB of video memory. While these boards capably render the flat 2D graphics of office applications and Web browsers, they shine when used for 3D gaming. The midrange boards will run you between $200 and $300. Many boards, such as those using NVidia's GeForce2 chip set, use faster Double Data Rate SDRAM memory, which doubles the speed by transferring data twice with every "tick" of the memory's clock, instead of just once.

The most advanced boards ship with a minimum of 64MB of onboard video memory, the most powerful GPU chip sets, and advanced 3D rendering capabilities. These powerhouse boards, capable of fill rates in the gigapixel-per-second range, sell for close to $400 and offer other advanced features such as hardware transform and lighting, which can speed up frame rates in games that support it, and full-scene antialiasing to clean up 3D images at low- and midrange resolutions.

All PCs ship with basic graphics capabilities: This can include a board based on an earlier generation chip set (such as a 3dfx Voodoo3, a Matrox Millennium G200, or an NVidia TNT2 GPU) or one of the budget-targeted boards (ATI's 3D Rage Pro, 3dfx's Voodoo4, or NVidia's GeForce2 MX), streamlined boards with the latest 3D graphics abilities but not all the bells and whistles.

The least desirable option for people who use 3D applications is an integrated graphics GPU built into the PC's motherboard. In an integrated graphics PC, you'll usually find a low-end graphics processor and a minimal amount of video memory (4MB to 8MB), some of which might be shared with normal system memory. This memory sharing allows PC makers to sell new PCs for as little as $300 but often leads to slowdowns when handling intensive graphics applications.

Other valuable features you might find in graphics boards include video input and output--a necessary addition for video editing and feeding to a television or VCR; DVD acceleration, t enhance DVD-video playback; DVI output, for the digital visual interface used by digital C displays and some flat-panel monitors; and multiple-monitor support.

The Future of 3D Graphics

In recent years, graphics board manufacturers have focused on 3D gaming as the primary driving force to develop new boards, according to Peter Glaskowsky, a senior analyst for 3D and multimedia at Cahners Microdesign Resources. But that will change. "Microsoft would like to integrate more 3D into the user interface, and in the next year or two they'll be in a position to make the whole Windows interface 3D. At that point, 3D will matter much more to the average user," he says.

Other options, such as integrated graphics, continue to grow because of their lower cost. (Jon Peddie Associates estimates that of the 160 million PCs that will ship this year, 100 million will use integrated graphics.) But most analysts feel that the 3D graphics board industry has a long life ahead of it.

"The ability to generate something that looks as good as *Toy Story* is only a few years away," says Glaskowsky. But duplicating photorealism in real time will probably take until 2010, he says. Perhaps another 10 to 20 years will pass before graphics boards emerge with power to make the trillions of floating point calculations required to replicate the subtlety of real life.

Related Topics: CRT, Graphics & Multimedia, Software, Technology